The Profitability of Broadway Shows

Emie Ung and Christine Zhou

25 November 2020

```
knitr::opts_chunk$set(echo = TRUE)
```

The Profitability of Broadway Shows

```
# Download  picture of Broadway
knitr::include_graphics("./Broadway_NY.jpg")
```


```
# Load packages we need
pacman::p_load(tidyverse, readxl, RColorBrewer, dplyr)
```

Description

Data Source: Broadway League (“https://www.broadwayleague.com/research/grosses-broadway-nyc/#weekly_grosses (https://www.broadwayleague.com/research/grosses-broadway-nyc/#weekly_grosses)”). This is the weekly grosses and associated data of Broadway shows in NYC filtered from between May 12, 2019 to Sept 1, 2019.

Exploring the Data

First, let's download the data.

```
#Loading excel file data
broadway_trend <- read_excel("./Broadway_Grosses.xlsx")
```

Summary of Columns in the Data Set

Then, we wanted to explore the data a little bit. We decided to look at some summary data to get a general overview of the profitability of Broadway shows.

```
summary(broadway_trend)
```

```
##     Week End                       Show               Type
##  Min.   :2019-05-12 00:00:00   Length:547         Length:547
##  1st Qu.:2019-06-02 00:00:00   Class :character   Class :character
##  Median :2019-06-30 00:00:00   Mode  :character   Mode  :character
##  Mean   :2019-07-02 05:08:00
##  3rd Qu.:2019-07-28 00:00:00
##  Max.   :2019-09-01 00:00:00
##
##    Theatre              #Prev            #Perf          Grosses
##  Length:547         Min.   :0.0000   Min.   :0.000   Min.   : 206060
##  Class :character   1st Qu.:0.0000   1st Qu.:8.000   1st Qu.: 610238
##  Mode  :character   Median :0.0000   Median :8.000   Median : 916256
##                     Mean   :0.1499   Mean   :7.686   Mean   :1047900
##                     3rd Qu.:0.0000   3rd Qu.:8.000   3rd Qu.:1311728
##                     Max.   :8.0000   Max.   :9.000   Max.   :3223611
##
##  Grosses\r\nPrev Week  gross_profit        Attend      Attend\r\nPrev Week
##  Length:547           Min.   :0.2000   Min.   : 1936   Min.   :  974
##  Class :character     1st Qu.:0.5800   1st Qu.: 6538   1st Qu.: 6510
##  Mode  :character     Median :0.7900   Median : 8173   Median : 8162
##                       Mean   :0.7769   Mean   : 8714   Mean   : 8703
##                       3rd Qu.:0.9450   3rd Qu.:11098   3rd Qu.:11144
##                       Max.   :1.4100   Max.   :16334   Max.   :16334
##                                                        NA's   :11
##       Cap
##  Min.   :0.3900
##  1st Qu.:0.8200
##  Median :0.9500
##  Mean   :0.9052
##  3rd Qu.:1.0000
##  Max.   :1.0400
##
```

We changed a few of the original column titles to the following:

"GG%GP" -> "gross_profit"

"% Cap" -> "Cap"

Key:

"#Prev" - Number of previews (pre-show performances)

"#Perf" - Number of performances

“GrossesWeek” - Grosses from previous week

“Attend” - Number of people who attended

“Cap” - Percent of seats filled

Structure of the Data Set

```
str(broadway_trend)
```

```
## tibble [547 x 12] (S3: tbl_df/tbl/data.frame)
##  $ Week End            : POSIXct[1:547], format: "2019-05-12" "2019-05-19" ...
##  $ Show                : chr [1:547] "AIN'T TOO PROUD" "AIN'T TOO PROUD" "AIN'T TOO PROUD"
"AIN'T TOO PROUD" ...
##  $ Type                : chr [1:547] "Musical" "Musical" "Musical" "Musical" ...
##  $ Theatre             : chr [1:547] "Imperial" "Imperial" "Imperial" "Imperial" ...
##  $ #Prev               : num [1:547] 0 0 0 0 0 0 0 0 0 0 ...
##  $ #Perf               : num [1:547] 8 8 8 8 8 8 8 8 8 7 ...
##  $ Grosses             : num [1:547] 1494459 1463394 1576625 1513346 1548022 ...
##  $ Grosses
## Prev Week: chr [1:547] "1445734" "1494459" "1463394" "1576625" ...
##  $ gross_profit        : num [1:547] 1.03 1.01 1.08 1.03 1.05 1.09 1.08 1.06 1.03 1.04 ...
##  $ Attend              : num [1:547] 11365 11373 11378 11357 11392 ...
##  $ Attend
## Prev Week : num [1:547] 11242 11365 11373 11378 11357 ...
##  $ Cap                 : num [1:547] 1 1 1 1 1 1 1 1 0.99 0.99 ...
```

Now we have a better sense of the types of data we’re working with.

Research Questions

1. What type of show (Play, Musical, or Special) was the best investment?

Type of show vs grosses - box plot

Average show gross profit vs type of show - density plot

2. How does the attendance of a show change with time?

Percentage capacity vs time - line graph

Which Type of Show is the Best Investment?

We now want to arrange the data to get concise information to see which type of show is the best investment in the theater. We want to get rid of shows that have no type, and arrange the rest of the data according to shows and their type.

```
# Create new data frame for sorted data
broadway_sorted <- broadway_trend %>%
  filter(`#Perf` > 0) %>%   # filter out shows with only previews, and no performances
  select(Type, Show, Grosses) %>%
  group_by(Type, Show) %>%
  summarize(total_shows = n(), mean = mean(Grosses)) %>%
  mutate(mean_gross = mean/1000000) %>%
  arrange(desc(mean))
```

```
## `summarise()` regrouping output by 'Type' (override with `.groups` argument)
```

```
# Print sorted data frame
broadway_sorted
```

```
## # A tibble: 46 x 5
## # Groups:   Type [3]
##    Type    Show                                    total_shows    mean mean_gross
##    <chr>   <chr>                                         <int>   <dbl>      <dbl>
##  1 Musical HAMILTON                                       17  3.05e6       3.05
##  2 Musical THE LION KING                                  17  2.34e6       2.34
##  3 Musical MOULIN ROUGE!                                   6  2.03e6       2.03
##  4 Special MORRISSEY                                       1  1.88e6       1.88
##  5 Play    TO KILL A MOCKINGBIRD                          17  1.85e6       1.85
##  6 Musical WICKED                                         17  1.78e6       1.78
##  7 Special DAVE CHAPPELLE                                  2  1.53e6       1.53
##  8 Musical ALADDIN                                        17  1.52e6       1.52
##  9 Musical AIN'T TOO PROUD                                17  1.51e6       1.51
## 10 Play    HARRY POTTER AND THE CURSED CHILD, PA~         17  1.45e6       1.45
## # ... with 36 more rows
```

This data tells us the individual show grosses, so we can find any outliers. For example, Hamilton's Gross revenue was $3,050,533.70, or over $3 million.

```
# Plotting Grosses for Types of Show in a Box Plot
  ggplot(data = broadway_sorted, mapping = aes(x = Type, y = mean_gross, fill = Type)) +
    scale_fill_brewer(palette = "Dark2") +
  geom_boxplot() +
  labs(title = "Figure 1: Summer 2019 Mean Broadway Grosses",
       x = "Type of Show",
       y = "Mean Gross Revenue (millions of dollars)")  +
theme(legend.title=element_blank())
```

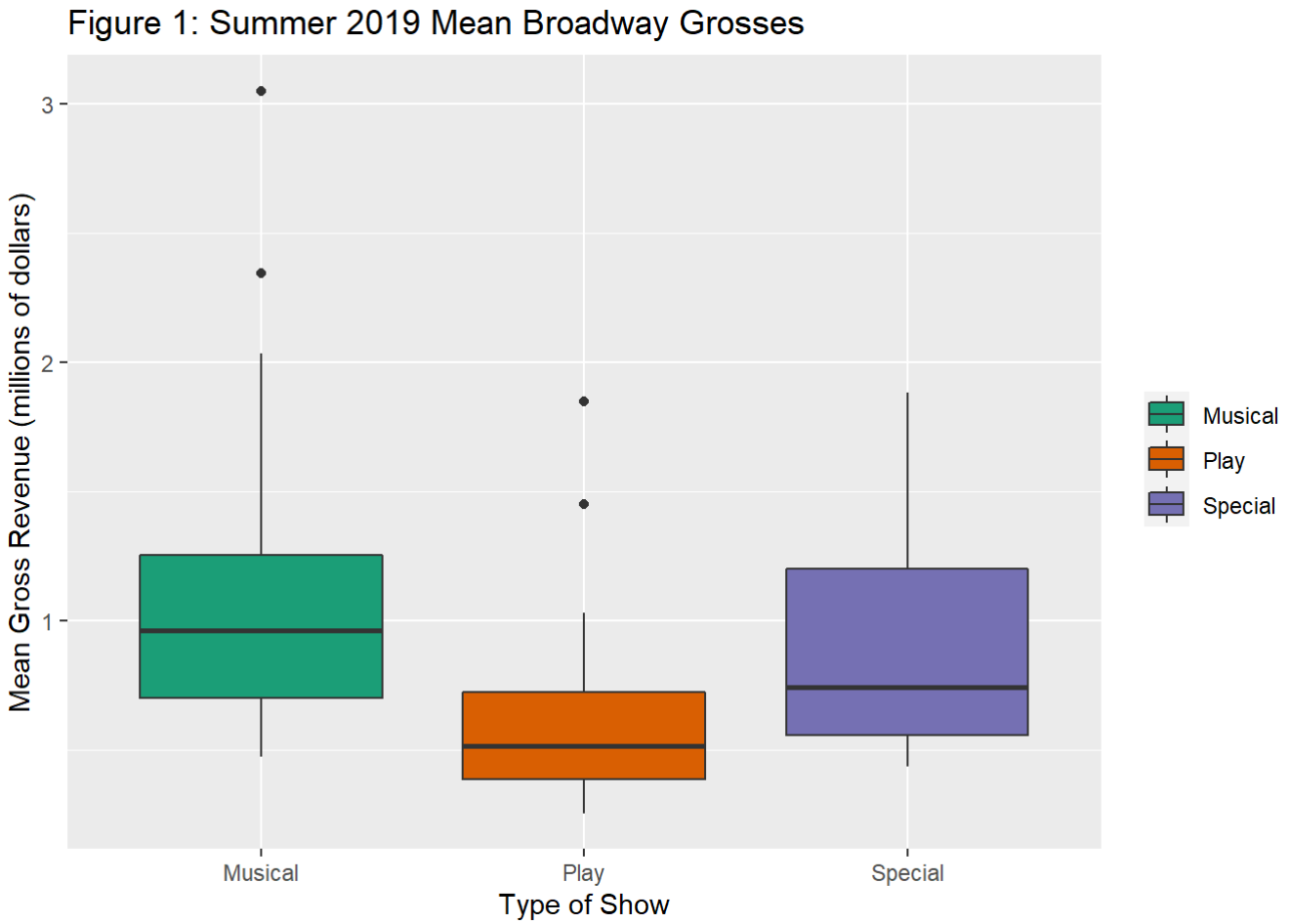

Figure 1 interpretation: From this figure we see that musicals have the largest range in terms of gross revenues of the three types of shows represented. While plays have a slightly larger range than specials, the average gross revenue is lower. One musical has a mean gross revenue greater than $3 million, while no specials or plays have revenue greater than $2 million. From this chart, musicals are the best investment, followed by specials and plays.

Note: We considered a dot plot, but the results were too spread out that we couldn't determine any real data for each variable. In the end, we chose the box plot instead.

```
# Plotting Gross Profit Ratio
broadway_trend %>%
group_by(Type, Show) %>%
select(Type, Show, gross_profit) %>%
summarize(mean_gross_profit_ratio = mean(gross_profit)) %>%
mutate(mean_gp_ratio = mean_gross_profit_ratio*100) %>%

ggplot() +
    geom_density(mapping=aes(x=mean_gp_ratio, color=Type), size=1) +
    scale_color_brewer(palette = "Dark2") +
labs(title = "Figure 2: Distribution of Gross Profit Ratio for Broadway Summer 2019",
     x = "Gross Profit Ratio (%)",
     y = "Proportion of Shows") +
theme(legend.title=element_blank())
```

```
## `summarise()` regrouping output by 'Type' (override with `.groups` argument)
```

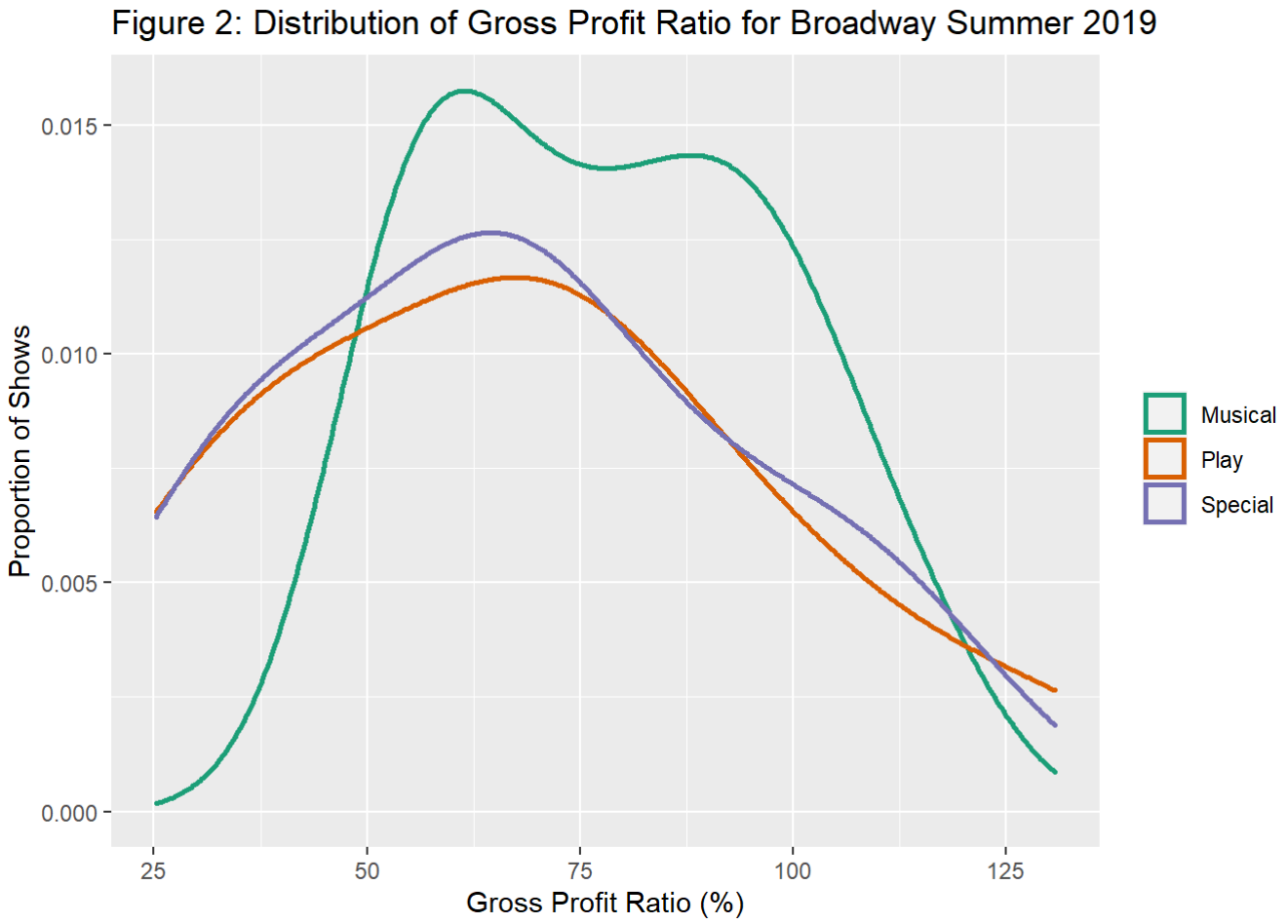

Figure 2 interpretation: The range of gross profit ratios is about 25% to 130%. It is more likely that a Broadway musical will have a gross profit ratio anywhere between 50-100%, while a greater proportion of specials and plays have gross profit ratios between 40-80%. Therefore, from this data, we can see that musicals are more profitable.

How Does the Attendance of a Show Change with Time?

Now we want to see if attendance for shows changed over the summer.

```
# Plotting Attendance Percentage Averages
broadway_trend %>%
select (Type, Show, Cap, `Week End`) %>%
mutate(percent_capacity = Cap*100) %>%

# Plot line graph
ggplot(mapping=aes(x=`Week End`, y=percent_capacity, color=Type, fill=Type)) +
  scale_color_brewer(palette = "Dark2") +
  scale_fill_brewer(palette = "Dark2") +
  geom_smooth(method = "loess") +
  labs(title = "Figure 3: Broadway Summer 2019 Attendance",
     x = "Date",
     y = "Capacity (%)") +
theme(legend.title=element_blank())
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Figure 3: Broadway Summer 2019 Attendance

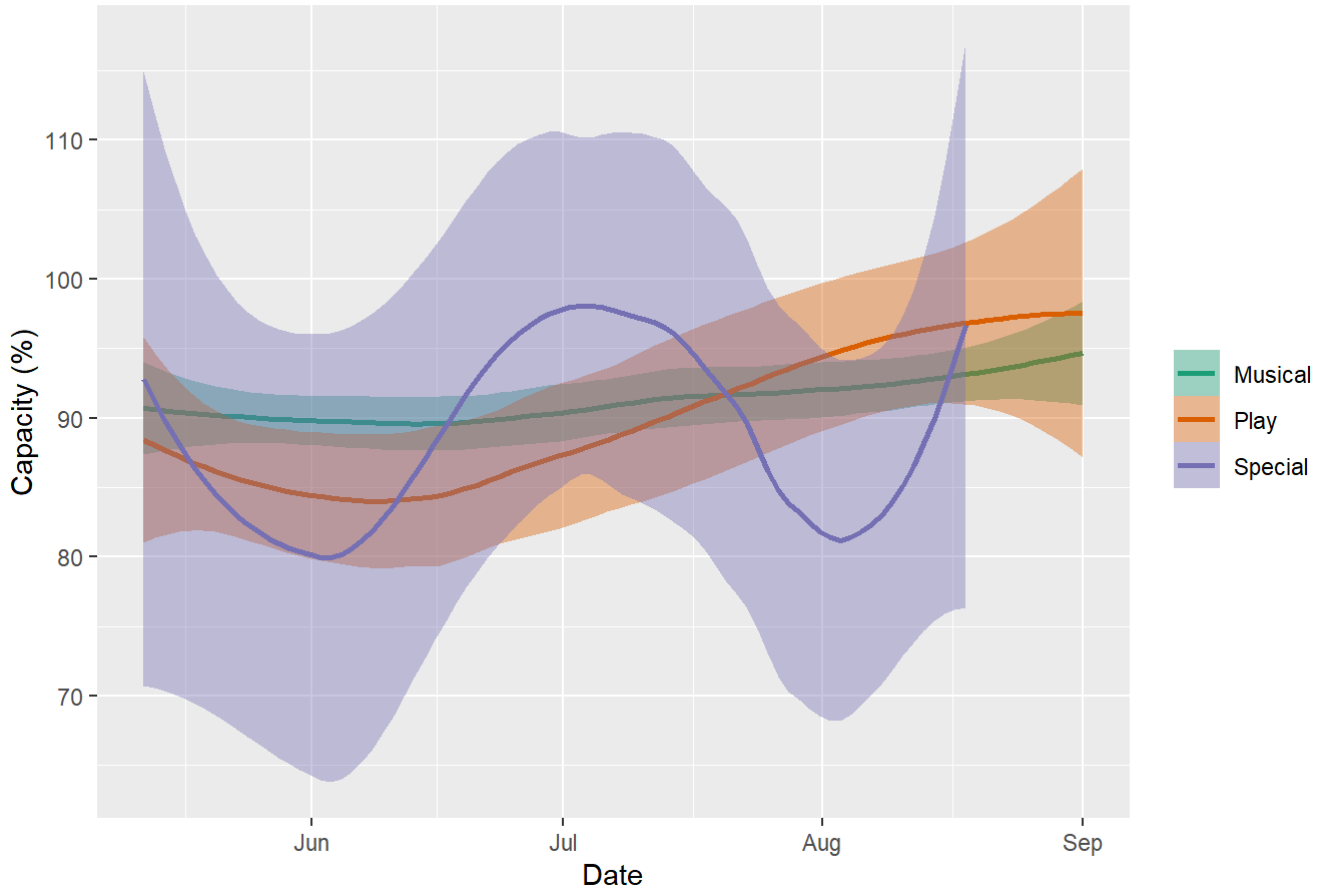

Figure 3 interpretation: We can see that the attendance for specials range much more than for musicals and plays. However, the average capacity for each date was all above 80%. There was a slight increase in percentage capacity for both musicals and plays through the months, but plays had a larger range of percent capacity than musicals.

Answers to Research Questions

1. What type of show (Play, Musical, or Special) was the best investment?

From the data in Figures 1 and 2, it appears that musicals are the best investment. The average gross is the highest for musicals, and the values of Q1 and Q3 are the highest. However, the outliers for gross revenue in musicals may be the cause of the higher values. Since a larger percentage of Broadway musicals seem to have higher gross profit ratios than plays and specials, they also seem to be a better investment in this respect. There are also the most data points for musicals within our data range, so this could be the reason for the slightly higher grosses and higher proportion of musicals with high gross profit ratios.

2. How does the attendance of a show change with time?

The general trend for all types of shows is a slight increase in attendance through the summer. However, the attendance for plays seems to have increased the most by percentage point. Attendance for specials fluctuated too much to determine whether there is actually an increase in average capacity but musicals also have a slight increase in percent capacity.